Submitted by: Terralogic, Inc.

VIRGINIA GIS REFERENCE BOOK

General Application Category: Health Department/Epidemiology

Product /Service/Function Name: Epidemiology

P/S/F/ Description: Application used by local governments and possibly others to track and analyze incidents of disease. The application would allow users to view and report local epidemiology incident information to assess patterns and the overall severity of existing conditions and/or threats. The epidemiology application will utilize database information collected at each Health Department office, and can be geographically related to an exact address, city, or a zip code. A user may use this application to assess the risk of contracting certain diseases, such as influenza, at their street or zip code level. Additionally, users may add adjacent locality data to further assess risk.

Product /Service/Function

1. Spatial Data

Spatial Data Definition: (ESRI, GIS Glossary, 1996) Information about the location and shape of, and relationships among geographic features, usually stored as coordinates and topology. In general terms, spatial data is geographic information.

Minimum Requirements:
The geographic or spatial data selected for this application can vary from freely available census data, to the utilization of accurate E-911 centerlines, derived from aerial photography. TIGER (US CENSUS) line or street files are the most inexpensive spatial data sets that contain address location information. This addressing information is tied to the road centerline, as an address range. Each road segment houses all addresses within one city block. If trying to locate address 555 on a street segment where the address ranges from 500 to 599, the location would be identified in the TIGER data as a point, located a little over halfway up the road, on the left.

Alternatively, the locality can develop its own centerline file from existing aerial photography. This however, is complicated and requires field verification of each address that is deemed inhabitable.

If the locality houses an E-911 database, the developer should consider utilizing this more accurate, verified information.

Optional Requirements:
Additional spatial data may include:

- Hospital point locations in GIS format (ESRI shape file format)
- Adjacent locality spatial data layers that may be available.
- Other spatial data layers that may be helpful in analysis (i.e., livestock, demographic, forestry, etc.)

2. Attribute Data

Attribute Data Definition: (ESRI, GIS Glossary, 1996) 1. A characteristic of a geographic feature described by numbers, characters, images and CAD drawings, typically stored in a tabular format and linked to the feature by a user-assigned identifier.

Minimum Requirements:
Typically, each locality is required to report specific infectious diseases to the Commonwealth of Virginia Department of Health. The department is divided into "Health Districts". The district offices house a Microsoft Access database, tracking each incident in their district. This information is eventually compiled at the state level and published in various forms. The locality wishing to develop an epidemiology application would benefit from obtaining this data at the district office level. The protocol for obtaining this local information is to submit a formal request for data through your local department

Attribute data will generally be delivered in two forms. One form will be tabular data in a ".dbf" file format (part of an ESRI shape file set). These spatial data are typically and best limited to unique identifier column or columns that hold pertinent spatial information. Additional attribute information should be delivered in a separate typical database structure (ASCII text file, spreadsheet, database). These data can contain all other information that needs to be tied to the spatial data via the unique identifier. All data structures and naming conventions should be in standard ANSI formats.

Optional Requirements:

- Live (Internet) connectivity to the State Department of Health database for current information on all localities. This live connection could also supply a means for reporting incident information to the proper authorities.

- Tabular data that are related to hospital points including disease specialties, capacity to deal with incidents, and number of beds. Records should be uniquely identified by hospital name or unique identifier.

3. Data Acquisition Options (integrated with VBMP digital orthos)

All spatial and attribute data sets are freely available from the local Health District or downloaded from the US Census Bureau.

The integration of these data with the VMBP digital orthophotographs will provide a highly accurate base map for better location or positional analysis. This may require adjusting, or conflating the existing spatial data to the new VBMP orthophotos.

4. Data Conflation Options (integrated with VBMP digital orthos)

Conflation is the method whereby a geographic feature is adjusted to fit a more accurate base map. Conflation will help to ensure that all locality data are positionally in the same space in relation to each other on earth. With the use of an accurate digital image, the position of associated vector data is conflated to overlay the position of the same features on the image base. Typically the best base map for conflation purposes is a current, high-resolution digital orthophotography product. It is paramount that the orthophotography is as accurate as possible since any error in the imagery will also be reflected in any feature that is located using that imagery. Feature layers that were created by on-screen digitizing directly from the orthophotography should not need conflation if the orthophotography is being used as the base.

The conflation process can occur in a variety of ways, with the least sophisticated being a "best-fit" methodology. The best-fit method is a visual inspection or comparison of a geographic feature's current position to where it is or should be located on the more accurate base map. This method would either entail:

1) moving the entire road lines layer across the imagery (like sliding one sheet of paper over another) until the greatest number of roads aligned as closely as possible with their counterparts on the imagery; or,
2)moving individual road lines, or sets of road lines, so that they align as closely as possible with their corresponding roads on the imagery.

These methods may be the best solutions in many cases since it will take less time than other options and will be the fastest to implement. Additionally, the availability of the VBMP orthophotography will provide a good base map in which to "fit" features. This method uses visual judgement to determine the best fit of the features.

Another conflation option includes rubber sheeting: a method using control points or existing boundaries to establish the new geographic position of a feature. Using roads as an example, control points at road intersections that can be clearly identified on both the digital imagery and on the road vectors are used to "stretch and shrink" the vector roads so that their positions correspond at the control points. The more control points that are used, the more precisely the data will fit. This method uses two sets of control points and a GIS algorithm to adjust the vector feature locations.

Finally, the most accurate method of conflating data includes the use of Global Positioning Satellite technology (GPS), or traditional survey instruments to accurately locate each desired object's physical location. While this is very accurate in most cases, the existence of an orthophotography base map product may be the best source for conflation because when viewing data/maps with different layers present, it is desirable to have the framework or vector data "fit" over the orthophotography. If a wide base of accurate GPS spatial data is already present then conflating the orthophotography may be satisfactory. This method uses direct field measurements, the most spatially accurate data as control points, and a GIS algorithm to adjust the imagery.

Any data not collected directly from the image base will likely need some level of conflation if it is to be overlaid effectively with the orthophotography.

5. GUI / Programming Options

A graphical user interface (GUI) enables a user to perform desired tasks by using a mouse to choose from a "dashboard" of options presented on the display screen. These are in the form of pictorial buttons (icons) and lists. Some GUI tools are dynamic and the user must manipulate a graphical object on the screen to invoke a function. For example: moving a slider bar to set a parameter value (e.g., setting the scale of a map). The GUI is the interface used to interact with the data and perform analysis functions.

An Epidemiology application should have the ability to show and thematically map diseases by specific areas of interest, such as census tracts or municipal boundaries, and report the descriptive attribute information to the user. Links to disease information with medical and technical descriptions should be provided, including how the area of interest compares with the other areas of interest. This type of application is generally used by the Department of Health and could optionally be made available to field personnel and the public. Some components of the interface may include:

1) Input text boxes that enable searches against a database of diseases can be used for both feature identification and for data input. There should be a one-to-one correspondence between the search components and the attributes in the database so that additional time does not have to be spent parsing the data before searching.
2) A summary of the input information should be easily visible from the map view port.
3) A map view port large enough for users to easily ascertain location information and object information.
4) If the system has additional layers (other than boundaries) available in the application such as hydrology, building footprints etc., then the system should give the user the ability to turn certain layers on and off for reference purposes. This usually appears as a scrolling list of layer names and check boxes.
5) The application interface should also give the user the ability to view current and historical information on diseases for a particular location and give the user the ability to run additional queries.

GIS software can be modified utilizing a variety of programming languages or scripting languages that may vary depending upon the system architecture. This will enable the customization of the application interface and functions to meet non-standard requirements. Languages such as Microsoft Visual Basic are commonly used to invoke macros and customized functions such as GIS queries. Commonly used languages include: Visual Basic, C++, Java, HTML, ASP, Cold Fusion, JSP, PERL, PHP and CGI.

For data that can be modeled as a "system" of interconnecting parts, certain GIS applications enable the development of "smart feature" models (called a GEO Database by ESRI). These allow feature characteristics to be defined within a database, enabling system constraints and functionality to be stored with the features. This reduces the amount of programming in the interface. An example would be a pipe inventory that presents the pipes and connectors of the system graphically. If a user tried to place an inappropriate fitting between two pipes (a fitting that connects a 6 inch pipe with an 8 inch pipe, would not be allowed between two 8 inch pipes) the system would flag the error and warn the user. This also helps when the system is being used for system maintenance.

6. Internet Functionality and Options

Internet functionality generally refers to public access to a system through the use of the Internet, or a public portal to a system that is password protected so that while the public can view the portal over the Internet, only authorized users can gain direct access to the application. These access types are generally enabled through a standard web browser. This is different than Intranet access, where access is limited to a specific computer network, usually one agency or group of agencies, and allows no public access.

An Epidemiology application is usually publicly accessible and may be available to the public through the Internet. For access by field personnel, or other agencies, presenting this information over the Internet may be advantageous. If public viewing of the system's feature data, usage data, or historical event data is desired, presenting it in a standard map window with related base data can be achieved fairly easily. These applications can take advantage of web browser interfaces, Internet transfer protocols and the flexibility of the Internet programming languages. Since web browser interfaces are so widely recognized and understood and since the programming languages are robust and enable the customization of the standard application interfaces, they are used in creating these types of applications. The advantage of Internet protocol interfaces is that data transfer is very fast.

Standard Internet mapping functionality would include basic GIS functions available in a thin client GIS application, such as ESRI's ArcExplorer (i.e. Zoom In, Zoom Out, Pan, Identify, Query, Thematic Mapping, etc.). Additional functionality may include appropriate hyperlinks to critical and related information on the Internet related to certain queries or operations within the application. An Internet application allows the organization to share its spatial and tabular information to all authorized users via a familiar Internet browser interface. This eliminates multiple software license fees. Additionally, the Map Server (Web Server) is the only GIS hardware/software component that would be managed by the localities Information Technology Department.

7. Minimum Technical Requirements
The basic technical requirements needed to set up the GIS component of an Epidemiology system are listed below.

1) A Basic working knowledge of a leading GIS software, and Internet browser are required.
2) A Pentium III or greater CPU, with a minimum of 128MB Ram, 16MB Video Card, is required. A high speed Internet connection is recommended for GIS Internet application deployment and analysis.
3) Most leading GIS software is customizable using MS Visual Basic or another common language. It is suggested that the developer have a working knowledge of (at least) Visual Basic before attempting GUI development.
    a. T1 or better connections to the Internet for access by field personnel.
    b. Server should be RAID level 5 with two-stage back-up (mirrored systems as well as tape back-up) to minimize data loss and to enable quick data recovery.

Optimum Technical Requirements:
For a fully integrated mapping and robust Epidemiology system, the options are near limitless. Below are some of the components that could be implemented to utilize the full benefit of a spatial/GIS based application.

1.) A local Epidemiologist may be utilized to better understand certain diseases and the risks associated with infection.
2.) A robust RDBMS networked to other departments as a central repository for all locality spatial data.
3.) An integrated work order system capable of tracking system conditions and managing work orders.
4.) Implementation of a 2-stage back-up and recovery system for rapid recovery of system failures.

In the case where a local government employs a highly capable Information Technology Department, other languages may be considered, such as JSP, Java, Visual Basic, ASP, and Cold Fusion. In most cases, these languages are related to Internet application development. A web developer with three years of experience should be able to customize and/or develop a unique Internet Map Server application.

8. Administrative / Management Requirements

Management concerns will involve technical support, system maintenance and, of course, human resource management issues of a technical product. These issues are minimized if the maintenance and/or hosting of the application are contracted to a GIS application development and hosting organization. Technical and administrative issues become more critical and consuming when developing and/or hosting an application in-house. General expertise in GIS is suggested if outsourcing application development and hosting. In-house application development and hosting will require GIS specialist human resources, advanced web programming human resources, and significant technical material resources (hardware/software).

Management of an epidemiology application should be concerned with data development, application development and system maintenance. The data required for this application typically exists at the local Health Department or the Census Bureau. The manager must be tasked with obtaining the necessary data from these respective organizations, and developing a repeatable process for updating/uploading the information to the specified application data source. In this case however, the Health Department Epidemiology database(s) are already in a compatible file format (Microsoft Access) with most industry standard GIS systems. The critical process is to link these tabular records to their corresponding geographic feature.

A manager or administrator implementing a project of this nature will need a strong project management skill set due to the variety of the components that will be involved and have (or develop) a thorough understanding of water systems and the role of water meters in that system. There are seven (7) main areas where administration and management requirements will need to be concentrated:

1.) Fiscal – Pre-planning and research on how other localities implemented similar systems is recommended. A manager with good fiscal, budgeting and money management skills will be helpful.
2.) Personnel - This type of implementation should require limited staff resources but some technical expertise in GIS and application development will be required. If contractors are hired for any part of the implementation or maintenance, the manager would oversee their performance as well. Good personnel management skills will be required.
3.) Technology – A manager with a good understanding of GIS is recommended, but this type of application does not require extensive technical knowledge. The manager will need experience in selecting technical human resources.
4.) Political/Stakeholder Involvement – The ability to work with a limited number of external stakeholders may be needed. In this case, the Service Authority or Public Works Department may want connectivity to the planning data so that new installations can be anticipated, or to the water systems data so that repair notices can be sent to the affected customers.

5.) Training – Training will likely be limited to the technical personnel running and maintaining the system and potentially the staff of other agencies.
6.) Operational – Day-to-day operations will be limited unless the locality is large and requires extensive daily maintenance. Scheduling may be required.
7.) Maintenance – Once the system is in place, on-going maintenance will need to occur. Before the system ever reaches this stage, the manager should be able to develop a plan identifying what resources are needed, what resources are expected to be available, and how maintenance will be performed. This will help guide the system requirements, and help ensure that an initial implementation is not undermined by the inability to support it over time.

In general, management concerns will involve technical support, system maintenance, and human resource management of technical staff. Technical and administrative issues become more critical and consuming when developing and/or hosting an application in-house. General expertise in GIS is suggested if outsourcing application development and hosting. In-house application development and hosting will require a GIS specialist, an advanced web programmer, and technical material resources (hardware/software).

9. Cost – Cost/Benefit

The cost of developing an Epidemiology application utilizing the local Health Department Database and freely available US CENSUS data varies based upon the amount of customization preferred/required. Customization is based on the needs of the user, and not every locality has identical needs.

The duration of the programming effort accounts for the majority of the cost of the application. In most cases, the application can be developed in under 160 hours.

Application Development: The cost of developing an epidemiology application (in-house) ranges from $7,500 to $15,000.
Programming the application, which includes a GUI that can post custom queries, accounts for at least 65% of the project costs.

Program Development: Developing an Epidemiology program from scratch will include application development, data production of areas of interest (boundaries), data conflation, hiring of staff, contract development/negotiations, multi-agency networking, database creation, field verification of data, all hardware and software, training, etc. This process could take a couple years and run into the hundreds of thousands of dollars. Additional maintenance costs would also need to be projected.

An epidemiology program, that includes application development, can help save money by:

- Enabling spatial analysis and understanding of which diseases pose a threat to specific areas of interest;

- Enabling the implementation of a system component maintenance program that can help avoid costly repairs by tracking the condition of the components;
- Enabling more efficient use of hospital space and immunizations;
- Enabling more efficient collection, management, and monitoring of disease incidents;
- Enabling the tracking of needed future repairs for budget projections; and,
- Enabling other programs to use this data with their application in order to see how their data interact and relate to the Epidemiology data. For example, the toxins found in discharge from a certain area may have a direct or indirect correlation with the health of the persons in that area.

10. Standards / Guidelines Summary

Disease analysis should consider information contained beyond each border; each locality should include datasets including the surrounding jurisdictions

All GIS or Spatial data should be delivered in a format and projection that matches or can be conflated to the VBMP ortho base map. The attribute, or tabular data, provided by the Department of Health is in MS Access format and must be imported or connected as the system database.

A weekly or monthly Epidemiology database/report can be obtained at the local Department of Health. A relationship should be established with the Department of Health to develop and manage a data request and transfer mechanism. It should be noted that the data typically reported and distributed by the local Department of Health is very current compared to the state office, which distributes data from the prior month or year.

Where applicable, industry standards for computer formats (including ANSI), Internet communications protocols (like TCP/IP), and other relevant technology standards should be used.

The Information Technology Resource Management Guideline, “Model Virginia Map Accuracy Standards”, COV ITRM Guideline 92-1, 3/20/92; is a State of Virginia standard defining a common recognized standard to guide the collection of data for all map scales; a method for verifying and interpreting the data collected and map products produced; and a method of labeling data and map products. For more information: www.vgin.state.va.us/documents/guidelines-standards/guidelines-standards.html

The Federal Geographic Data Committee (FGDC), with consensus from local, state, federal, and private reviewers, created and maintain the “Content Standards for Digital Geospatial Metadata” (soon to become an international standard) to enable consistent and comprehensive recording of the content, quality, condition and other characteristics of spatial data. For more information: www.fgdc.gov

Seven (7) base map layers that have been recognized by the Federal Geographic Data Committee (FGDC), with consensus from local, state, federal, and private reviewers, are termed “Framework” layers. These layers are generally considered the main layers that most mapping organizations need to best enable and support their functions. They are: geodetic control, orthoimagery, elevation, transportation, hydrography, governmental units, and cadastral information. The FGDC has developed procedures, technology and guidelines (including basic attribute requirements) that provide for the integration, sharing, and use of these data and have also identified institutional relationships and business practices that encourage the maintenance and use of these data. For more information: www.fgdc.gov/framework

11. Startup Procedures/Steps

The need for the development of this type of application usually stems from an identified need for the service. Once this is recognized, a cost/benefit analysis will help determine if the resources and the benefits further warrant the development. If so, than the system itself can be addressed.

Regardless of whether the work will be performed in-house, or fully/partially contracted, a three-phase approach works well.

1) Needs Analysis - The first step is to do a “**needs analysis**” on the current system, or planned system. This will clearly identify and record the goals and resource needs of the project and will later be used to define the project steps. A needs analysis should include:
   - The overall goals and expectations of the system;
   - An inventory of current and expected resources;
   - Identification of components that need to be developed/added;
   - What types of products/documentation are expected;
   - What cooperative efforts and stakeholders will be involved;
   - What time parameters are involved;
   - What standards need to be used and what thresholds monitored; and,
   - How will the system be managed over time.

The needs analysis does not focus on implementation strategies, only on the what the system needs to do and what resources will be need to be involved.

2) With the needs analysis to guide and set goals, a **“systems design”** that meets those needs can be developed. This focuses on how the system gets built and determines:
   - What physical resources (specific hardware, software, etc.) will be used;
   - Who will manage what components of the system;
   - Where will the system reside;
   - Who will build/manage/maintain each part of the system;
   - How will the system be used, step-by-step, to achieve the goals; and,
   - What type of specific, on-going support will be established?

This focuses on the specifics of the system (the type, name, and size of computer), the human resources to manage it, and the way it will work and be maintained. The particulars of any given aspect may evolve as the system design is developed and specific questions or hurdles are discovered.

3) Now that an actual design has been determined, an **"implementation plan"** describing how the system will be implemented needs to be developed. This begins the physical implementation of the system. The implementation plan defines the:
   - Order of the implementation steps, (putting data in the system relies on the existence of the data);
   - Time, money (costs), personnel resources, and stakeholder dependencies (this must occur in order for that to occur);
   - Deliverables, formats, and documents by tasked entity (who does what, when and how);
   - Implementation phases and task timelines (this task should take this long);
   - Roles of the stakeholders, their expected tasks and commitments; and,
   - Future implementation/maintenance tasks and how they get accomplished.

If the work is be contracted, than these documents should outline the role(s) of the contractor(s) as well. This is the step-by-step guide to "making the system".

Some specific project steps may include:

1.) Researching standards, available data, other implementations, and possible outside funding sources.

2.) Inventory existing/expected resources (hardware, software, staff, money, etc).

3.) Develop an application outline/blueprint, focusing on the application's purpose, interface design, functionality, querying capabilities, and "look and feel". Stakeholders should be involved in this step.

4.) The attribute data will need to be obtained from the various sources mentioned earlier and normalized and related where necessary. Spatial data will need to be compiled from a variety of sources, or, if it is not available, then it will need to be collected and developed.

5.) Determine the entity/entities that will be performing data development functions, application development functions and application hosting functions and create a project plan with budget numbers.

6.) Develop an implementation plan that includes timelines and milestones.

7.) Create a data development/transformation plan that includes metadata definitions, a database schema, and data dictionaries with relational information.

8.) Readdress your project plan, timelines and budgets as a final initial process before committing resources.

It is recommended that the database application functions be addressed and implemented before the mapping functions.

12. Estimated Time Line and/or Implementation (stand alone) schedule

Time line and implementation schedules will be determined on an individual entity/locality basis since there is a very wide variety of implementation approaches based on the current status of data and other resources of a particular locality.

The estimated time to develop this application varies based on functionality. This can be as little as one month, to as much as a few months. Typically this type of application can be developed in approximately 150 hours.

If the locality is planning a comprehensive overhaul or starting from the ground level, the data collection process, system implementation and integration may take a couple of years. Hiring staff, manually compiling data and digitizing it, researching information from long-time Health Department or other staff, implementing a system, programming the application, etc., will all contribute to the time estimates. These types of jobs can typically be performed in less time, and possibly less overall money, if they are contracted to qualified firms since these firms have dedicated and knowledgeable staff experienced and ready to go.

13. Best Practice Examples in Virginia

The best practice example of an epidemiological application in Virginia can be located at: http://vadiseases.gisbrowser.com/. Although relatively simple, it provides the user with an easy mechanism to report disease information with little or no GIS experience. This is a State Epidemiology application example and does not go into the spatial or tabular detail that can be granted by utilizing the local Health Department's data. The application using the complete local Health Department database could locate the exact address of each incident of disease.